Controlled Mode Distinguishability for Cybersecurity

Dawei Sun, Inseok Hwang, *Member, IEEE*, and Martin Corless, *Life Member, IEEE*

***Abstract*—Cyber-physical systems (CPSs) are a class of systems integrating cyber and physical components, and their security issues have gained a lot of attention in recent years. CPSs are modeled as hybrid systems in this letter since the logical and physical behaviors of CPS can be mapped to the discrete-state and continuous-state dynamics of the hybrid system, respectively. Motivated by the importance of situation awareness in an adversarial environment, we consider the mode distinguishability problem for a class of hybrid systems that can describe compromised CPSs. It is found that even though some modes of the hybrid system may not be distinguishable without knowing the attack inputs, the modes could be controlled distinguishable, which means their behaviors can be differentiated under certain control inputs. In this letter, the characterization of controlled distinguishability is studied, and the problem of finding control inputs for mode identification is proposed.**

***Index Terms*—Hybrid systems, switched systems.**

I. Introduction

NOWADAYS, critical systems are increasingly interwoven with cyber components for high-level automation and intelligence. However, the close interaction between the physical process and the cyber components makes cyber-physical systems (CPSs) vulnerable to cyber-attacks, which has been brought to the attention of researchers [1], [2], [3]. A CPS commonly has multiple modes of operation such that internal and/or external variations can be accommodated [4], and thus the behavior of a CPS is not only governed by its physical dynamics but also the logic command that decides the mode of operation. Therefore, a hybrid systems approach is a powerful tool for modeling and analyzing a CPS [5] as the cyber-physical coupling of CPS can be represented by the interaction between the discrete-state and continuous-state dynamics of the hybrid system.

In the hybrid system framework, both the switching attack that maliciously alters the operational mode and the data injection attack that falsifies the control signal can be considered simultaneously for cybersecurity. Specifically, the compromised CPS under the joint attacks can be modeled as a class of hybrid systems called hidden mode switched linear systems with unknown inputs [6] where the tampered logic behavior of the system is described by the unknown adversary mode switchings and the data injection attack is modeled as the unknown additive input. Then, the issues of attack containment and situation awareness for the compromised CPS can be studied through the problems of resilient stabilization and mode (discrete state) identification of the hybrid system, respectively. It should be remarked that situation awareness is vitally important for system operators to take effective strategy for defense, and it has been justified that system operators can attain significant advantages to stabilize the system under attack if they are aware of which mode is triggered [7], [8]. Although there are several works on designing the mode identification algorithm for the switched system subject to unknown attack inputs [6], [9], [10], the basic mode distinguishability for a switched system with unknown inputs has rarely been discussed extensively.

Mode distinguishability of a switched system describes whether the discrete state of the switched system can be recovered from the behavior of the continuous dynamics [11], [12]. Mode distinguishability for continuous-time switched systems with unknown inputs has been discussed in [13], which is further extended in [14] by considering a class of attack inputs. Besides, the invertibility of continuous-time switched systems with unknown inputs, which describes whether the discrete state and the unknown input can be simultaneously estimated, has been studied in [15]. It should be remarked that the requirement of mode distinguishability for some switched systems could be restrictive, which induces the concept of controlled mode distinguishability [16], [17]: whether some special inputs can be designed and applied to the switched system such that the behaviors of different modes are differentiated. We would like to note that the existing research has not studied controlled mode discernibility for switched systems with unknown attack inputs, which is the problem considered in this letter.

In this letter, we provide an alternative approach to characterize mode distinguishability and extend the results to controlled mode distinguishability, with application to CPS security analysis. Based on the analysis, a quadratic gaming approach is proposed to design the control input for mode identification. Our main contributions are: (1) complementing the analysis of mode distinguishability for discrete-time switched linear systems with unknown inputs;

Manuscript received March 2, 2021; revised May 7, 2021; accepted May 31, 2021. Date of publication June 10, 2021; date of current version June 30, 2021. Recommended by Senior Editor J. Daafouz. *(Corresponding author: Dawei Sun.)*

The authors are with the School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN 47906 USA (e-mail: sun289@purdue.edu; ihwang@purdue.edu; corless@purdue.edu).

Digital Object Identifier 10.1109/LCSYS.2021.3088301

(2) characterizing controlled mode distinguishability within the context of CPS security; and (3) formulating the mode identification problem for a controlled distinguishable switched system as a quadratic gaming problem.

The rest of the paper is organized as follows. In Section II, we introduce a switched system model that can describe a CPS subject to cyber-attack, and we define mode distinguishability and controlled mode distinguishability for the given system. In Section III, the characterization of mode distinguishability and controlled mode distinguishability is presented, and a quadratic gaming problem is associated with controlled mode distinguishability. In Section IV, an illustrative example is given. Finally, Section V concludes this letter.

II. Problem Formulation

Throughout this letter, we consider the following hidden mode discrete-time switched linear system as a CPS subject to attacks:

$$x_{k+1} = A_{q_k} x_k + B^c_{q_k} u^c_k + B^a_{q_k} u^a_k, \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the compromised continuous state (state), $q_k \in Q = \{1, 2, \ldots, M\}$ is the tampered discrete state (mode), $u^c_k \in \mathbb{R}^m$ and $u^a_k \in \mathbb{R}^{m_a}$ denote the control input and attack input, respectively, and A_{q_k}, $B^c_{q_k}$, $B^a_{q_k}$ are the system matrices of the subsystem corresponding to the mode q_k. Without loss of generality, we assume the continuous state x_k and control input u^c_k are available for mode identification, but the tampered mode q_k and attack u^a_k are unknown. For the tampered mode q_k, we assume it satisfies the dwell-time condition that the difference between any two switching instances should be sufficiently larger than dimension of the state space, motivated by the fact that extremely frequent mode switchings are typically prohibited for practical reasons.

For the i-th subsystem of system (1), the linear dynamics can be equivalently written in the following compact form:

$$\begin{aligned} X_{1:K} &= \mathcal{X}_i(K, x_0, U^c_{0:K-1}, U^a_{0:K-1}) \\ &\triangleq \mathcal{O}_i(K) x_0 + \Phi^c_i(K) U^c_{0:K-1} + \Phi^a_i(K) U^a_{0:K-1}, \end{aligned} \tag{2}$$

where $X_{1:K}$, $U^c_{0:K-1}$, $U^a_{0:K-1}$ are compact representations of the trajectory of the continuous state, the control input sequence, and the attack input sequence, that is,

$$\begin{aligned} X_{1:K} &= \begin{bmatrix} x_1^T & x_2^T & \cdots & x_K^T \end{bmatrix}^T, \\ U^c_{0:K-1} &= \begin{bmatrix} u_0^{cT} & u_1^{cT} & \cdots & u_{K-1}^{cT} \end{bmatrix}^T, \\ U^a_{0:K-1} &= \begin{bmatrix} u_0^{aT} & u_1^{aT} & \cdots & u_{K-1}^{aT} \end{bmatrix}^T, \end{aligned} \tag{3}$$

and $\mathcal{O}_i(K)$, $\Phi^c_i(K)$ and $\Phi^a_i(K)$ are defined as:

$$\begin{aligned} \mathcal{O}_i(K) &= \begin{bmatrix} A_i^T & A_i^{2T} & \cdots & A_i^{KT} \end{bmatrix}^T, \\ \Phi^c_i(K) &= \begin{bmatrix} B^c_i & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_i^{K-1} B^c_i & A_i^{K-2} B^c_i & \cdots & B^c_i \end{bmatrix}, \\ \Phi^a_i(K) &= \begin{bmatrix} B^a_i & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_i^{K-1} B^a_i & A_i^{K-2} B^a_i & \cdots & B^a_i \end{bmatrix}. \end{aligned} \tag{4}$$

For convenience, we will also use the following notation:

$$\begin{aligned} x_K &= \chi_i(K, x_0, U^c_{0:K-1}, U^a_{0:K-1}) \\ &\triangleq A_i^K x_0 + \mathcal{C}^c_i(K) U^c_{0:K-1} + \mathcal{C}^a_i U^a_{0:K-1}, \end{aligned} \tag{5}$$

where $\mathcal{C}^c_i(K)$ and $\mathcal{C}^a_i(K)$ are the last n rows of $\Phi^c_i(K)$ and $\Phi^a_i(K)$, respectively.

Now we provide the following definition of mode distinguishability with unknown inputs.

Definition 1 (Distinguishability): Modes i and j for system (1) are *distinguishable within K time steps* (*K-step distinguishable*) if for any $x_0 \neq 0$ and for any attack input sequences $U^i_{0:K-1}$ and $U^j_{0:K-1}$,

$$\chi_i(k, x_0, 0, U^i_{0:k-1}) \neq \chi_j(k, x_0, 0, U^j_{0:k-1}), \tag{6}$$

for some $k \leq K$. Modes i and j are *distinguishable* if for any $x_0 \neq 0$ and any attack input sequences $U^i_{0:\infty}$ and $U^j_{0:\infty}$ where

$$U^q_{0:\infty} = \begin{bmatrix} u_0^{qT} & u_1^{qT} & \ldots \end{bmatrix}^T \quad q = i, j \tag{7}$$

there is a $k \geq 1$ such that (6) holds. Modes i and j are called *indistinguishable* if they are not distinguishable.

Thus, two modes are K-step distinguishable if the two subsystems cannot generate the same trajectory within K time steps, assuming the initial state is nonzero and no control input is applied. For nonzero initial state and zero control input, K-step distinguishability is necessary and sufficient to distinguish two modes within K time steps without knowing the attack inputs.

Two modes could be indistinguishable if the switched system has some special structure. In these cases, we are interested in whether there exist some control inputs that can make the two modes behave in a distinguishable manner. Such inputs are called discerning inputs in [17] and [16]. The following definition formalizes the idea.

Definition 2 (Controlled Distinguishability): For system (1), modes i and j are *controlled distinguishable within K time steps* (*controlled K-step distinguishable*) if for each x_0, there exists a control input sequence $U^c_{0:K-1}$ such that for any attack input sequences $U^i_{0:K-1}$ and $U^j_{0:K-1}$, there is a $k \leq K$ such that

$$\chi_i(k, x_0, U^c_{0:k-1}, U^i_{0:k-1}) \neq \chi_j(k, x_0, U^c_{0:k-1}, U^j_{0:k-1}). \tag{8}$$

Modes i and j are *controlled distinguishable* if for each x_0, there exists a control input sequence $U^c_{0:\infty}$ such that for any attack input sequences $U^i_{0:\infty}$ and $U^j_{0:\infty}$, there is a $k \geq 1$ such that (8) holds. A control input which achieves distinguishability is called a *discerning input*.

With the above definitions, we will solve the following problems in this letter.

Problem 1: Characterize controlled distinguishability;

Problem 2: Given two modes that are controlled distinguishable, design a discerning input to distinguish the two modes and determine the mode.

III. Main Results

In this section, we first discuss a characterization of mode distinguishability with unknown input, and then solve the proposed problems.

A. Characterization of Indistinguishable States

In existing research, an augmented system approach is commonly considered for the characterization of mode distinguishability [11], [12], [13], [14]. Instead, we present an alternative approach for studying mode distinguishability with unknown input in a discrete-time setup, which can provide novel insights for understanding controlled mode distinguishability. We start with the following definition.

Definition 3 (Indistinguishable State): A state x_0 is a *K-step indistinguishable state* for modes i and j of system (1) if there exist $U^i_{0:K-1}$ and $U^j_{0:K-1}$ such that

$$\chi_i(k, x_0, 0, U^i_{0:k-1}) = \chi_j(k, x_0, 0, U^j_{0:k-1}) \tag{9}$$

for $1 \leq k \leq K$. A state x_0 is *indistinguishable* if there are $U^i_{0:\infty}$ and $U^j_{0:\infty}$ such that (9) holds for all $k \geq 1$.

Recalling (2), the collection of K-step indistinguishable states for modes i and j, denoted by M^K_{ij}, can be written as:

$$M^K_{ij} = \left\{ x_0 \in \mathbb{R}^n \middle| \Delta\mathcal{O}_{ij}(K)x_0 \in Im\{\Phi^a_{ij}(K)\} \right\} \tag{10}$$

where

$$\Delta\mathcal{O}_{ij}(K) \triangleq \mathcal{O}_i(K) - \mathcal{O}_j(K), \quad \Phi^a_{ij}(K) \triangleq [-\Phi^a_i(K) \;\; \Phi^a_j(K)]$$

and $Im\{\cdot\}$ denotes the column space. Immediately, we obtain the following lemmas.

Lemma 1: For all positive integers K, $M^K_{ij} \supseteq M^{K+1}_{ij}$.

Lemma 2: For all positive integers K, M^K_{ij} is a subspace of $\mathbb{R}^n$. Also there is a $\bar{K}$ such that $M^K_{ij} = M^{\bar{K}}_{ij}$ for $K \geq \bar{K}$.

For any $x_0 \in \mathbb{R}^n$, let

$$\begin{aligned} R_q(x_0) &\triangleq \left\{ \chi_q(1, x_0, 0, [u^a]) | u^a \in \mathbb{R}^{m_a} \right\} \\ R_{ij}(x_0) &\triangleq R_i(x_0) \cap R_j(x_0), \end{aligned} \tag{11}$$

Thus, $R_q(x_0)$ is the one-step reachable set (by the attacker) from x_0 for mode q, and $R_{ij}(x_0)$ is the one-step reachable set from x_0 subject to the attacks that prevent mode identification within one step (we use $[u^a]$ to denote the sequence with one term in (11) for consistency). Note that $R_{ij}(x_0)$ is nonempty if and only if $x_0 \in M^1_{ij}$. The following lemma characterizes $K+1$-step indistinguishable states.

Lemma 3: For any positive integer K, $x_0 \in M^{K+1}_{ij}$ if and only if

$$R_{ij}(x_0) \cap M^K_{ij} \neq \emptyset. \tag{12}$$

Proof: By definition, $x_0 \in M^{K+1}_{ij}$ if there are sequences $U^i_{0:K}$, $U^j_{0:K}$ such that

$$\mathcal{X}_i(K+1, x_0, 0, U^i_{0:K}) = \mathcal{X}_j(K+1, x_0, 0, U^j_{0:K}). \tag{13}$$

If $U^i_{0:K} = [{u^i_0}^T \;\; {U^i_{1:K}}^T]^T$ and $U^j_{0:K} = [{u^j_0}^T \;\; {U^j_{1:K}}^T]^T$ this is equivalent to

$$\mathcal{X}_i(1, x_0, 0, u^i_0) = \mathcal{X}_j(1, x_0, 0, u^j_0) =: x_1 \tag{14}$$

$$\mathcal{X}_i(K, x_1, 0, U^i_{1:K}) = \mathcal{X}_j(K, x_1, 0, U^j_{1:K}) \tag{15}$$

that is, x_1 is in both $R_{ij}(x_0)$ and M^K_{ij}. ■

Let M_{ij} denote the collection of all indistinguishable states. We call M_{ij} the *indistinguishable set.* Clearly, if $x_0 \in M_{ij}$, then there are sequences $U^i_{0:\infty}$ and $U^j_{0:\infty}$ such that

$$\chi_i(k, x_0, 0, U^i_{0:k-1}) = \chi_j(k, x_0, 0, U^j_{0:k-1}) \in M_{ij} \tag{16}$$

for $k = 1, 2, \ldots$

Theorem 1: A set $\mathcal{V} \subset \mathbb{R}^n$ satisfies

$$R_{ij}(x_0) \cap \mathcal{V} \neq \emptyset \quad \forall x_0 \in \mathcal{V} \tag{17}$$

if and only if for each $x_0 \in \mathcal{V}$ there are sequences $U^i_{0:\infty}$ and $U^j_{0:\infty}$ such that

$$\chi_i(k, x_0, 0, U^i_{0:k-1}) = \chi_j(k, x_0, 0, U^j_{0:k-1}) \in \mathcal{V} \tag{18}$$

for $k = 1, 2, \ldots$

Proof: To demonstrate the "only if" part, suppose (17) holds and consider any x_0 in $\mathcal{V}$. Property (17) implies that there are $U^i_{0:0}$ and $U^j_{0:0}$ such that (18) holds for $k = 1$. Now suppose that for some $K \geq 1$, there are sequences $U^i_{0:K-1}$ and $U^j_{0:K-1}$ satisfying (18) for $k = 1, 2, \ldots, K$ and let

$$x_K \triangleq \chi_i(K, x_0, 0, U^i_{0:K-1}) = \chi_j(K, x_0, 0, U^j_{0:K-1}) \in \mathcal{V}.$$

Since $x_K \in \mathcal{V}$, we must have $R_{ij}(x_K) \cap \mathcal{V} \neq \emptyset$. Therefore, there are inputs u^i_K and u^j_K such that

$$\chi_i(1, x_K, 0, u^i_K) = \chi_j(1, x_K, 0, u^j_K) \in \mathcal{V}.$$

With $U^i_{0:K} = [{U^i_{0:K-1}}^T \;\; {u^i_K}^T]^T$, $U^j_{0:K} = [{U^j_{0:K-1}}^T \;\; {u^j_K}^T]^T$, we obtain that

$$\chi_i(K+1, x_0, 0, U^i_{0:K}) = \chi_j(K+1, x_0, 0, U^j_{0:K}) \in \mathcal{V},$$

that is, (18) holds for $k = K+1$. By induction we can obtain infinite sequences $U^i_{0:\infty}$ and $U^j_{0:\infty}$ such that (18) holds for $k \geq 1$. To demonstrate the "if" part, consider any $x_0 \in \mathcal{V}$. Since there are sequences $U^i_{0:\infty}$ and $U^j_{0:\infty}$ such that (18) holds for $k \geq 1$, it follows that

$$x_1 \coloneqq \chi_i(1, x_0, 0, U^i_{0:0}) = \chi_j(1, x_0, 0, U^j_{0:0}) \in \mathcal{V}$$

Hence, $x_1 \in R_{ij}(x_0) \cap \mathcal{V}$ and $R_{ij}(x_0) \cap \mathcal{V}$ is non-empty. ■

The following result provides an important characterization of the indistinguishable set.

Corollary 1:

$$R_{ij}(x_0) \cap M_{ij} \neq \emptyset \quad \forall x_0 \in M_{ij} \tag{19}$$

and if $\mathcal{V}$ has property (17) then $\mathcal{V} \subseteq M_{ij}$. Hence, M_{ij} is the largest set with property (17).

Corollary 2: Suppose $M^K_{ij} \neq M_{ij}$ for a positive integer K, then

$$M^{K+1}_{ij} \neq M^K_{ij}. \tag{20}$$

Proof: Since $M^K_{ij} \neq M_{ij}$, M_{ij} does not contain M^K_{ij} and it follows from Theorem 1 that (17) does not hold for $\mathcal{V} = M^K_{ij}$. Therefore, there is an $x^*_0 \in M^K_{ij}$ such that $R_{ij}(x^*_0) \cap M^K_{ij} = \emptyset$. Lemma 3 tells us that $x^*_0 \notin M^{K+1}_{ij}$. Hence $M^{K+1}_{ij} \neq M^K_{ij}$. ■

Remark 1: Note that Lemma 3, Theorem 1, and Corollarys 1, 2 do not depend on the linearity of the subsystems.

Corollary 3: Suppose $M^K_{ij} \neq M_{ij}$ for a positive integer K, then $\dim\{M^{K+1}_{ij}\} < \dim\{M^K_{ij}\}$.

Proof: Since both M_{ij}^K and M_{ij}^{K+1} are subspaces and $M_{ij}^{K+1} \subset M_{ij}^K$ but $M_{ij}^{K+1} \neq M_{ij}^K$ it follows that $\dim\{M_{ij}^{K+1}\} < \dim\{M_{ij}^K\}$. ∎

From Corollary 3, we can conclude that n, the dimension of the state space, is an upper bound for the number of steps to distinguish two modes, if the initial state is not in the indistinguishable subspace. In addition, it is implied that the indistinguishable subspace M_{ij} is equal to M_{ij}^n, which can be explicitly computed using (10).

Corollary 4: Modes i and j of system (1) are K-step distinguishable for some positive integer K if and only if they are n-step distinguishable, where n is the dimension of the state space.

Consider any x_0 in M_{ij}^1. There is a pair of attack inputs that make the resulting trajectories indistinguishable within one step, i.e., there exist attack inputs $u_0^i(x_0), u_0^j(x_0)$ such that

$$(A_i - A_j)x_0 = [-B_i^a \quad B_j^a][u_0^i(x_0)^T \quad u_0^j(x_0)^T]^T. \tag{21}$$

The collection of pairs satisfying (21) is the nonempty affine subspace:

$$U_{ij}^a(x_0) = [u_0^i(x_0)^T \quad u_0^j(x_0)^T]^T + ker\left\{[-B_i^a \quad B_j^a]\right\} \tag{22}$$

where $ker\{\cdot\}$ denotes the nullspace and $[u_0^i(x_0)^T \quad u_0^j(x_0)^T]^T$ is a particular solution to (21). From (21) we can show that there exist matrices F_i, F_j such that

$$u^i(x_0) = F_i x_0, \qquad u^j(x_0) = F_j x_0, \qquad \forall x_0 \in M_{ij}^1. \tag{23}$$

Also there exist matrices E_i, E_j such that

$$Im\left\{[E_i^T \quad E_j^T]^T\right\} = ker\left\{[-B_i^a \quad B_j^a]\right\}. \tag{24}$$

Then, for any $x_0 \in M_{ij}^1$, $U_{ij}^a(x_0)$ can be expressed as:

$$U_{ij}^a(x_0) = [F_i^T \quad F_j^T]^T x_0 + Im\left\{[E_i^T \quad E_j^T]^T\right\}, \tag{25}$$

which implies that, for any $x_0 \in M_{ij}^1$,

$$\begin{aligned} R_{ij}(x_0) &= (A_i + B_i^a F_i)x_0 + Im\{B_i^a E_i\} \\ &= (A_j + B_j^a F_j)x_0 + Im\left\{B_j^a E_j\right\}. \end{aligned} \tag{26}$$

It now follows that a subspace $\mathcal{V}$ has the property that $R_{ij}(x_0) \cap \mathcal{V} \neq \emptyset \ \ \forall x_0 \in \mathcal{V}$ if and only if $\mathcal{V}$ is a $(A_i + B_i^a F_i, B_i^a E_i)$-controlled invariant subspace contained in M_{ij}^1 (see [18]). Hence we obtain the following result.

Theorem 2: For modes i and j of system (1), the indistinguishable subspace is the largest $(A_i + B_i^a F_i, B_i^a E_i)$-controlled invariant subspace contained in M_{ij}^1, or equivalently the largest $(A_j + B_j^a F_j, B_j^a E_j)$-controlled invariant subspace contained in M_{ij}^1, where (F_i, F_j, E_i, E_j) are obtained from (23) and (24).

Remark 2: Note that the $(A_i + B_i^a F_i, B_i^a E_i)$-controlled invariance property is a geometric property that is independent of the choice of F_i or E_i: if both (F_i, F_j, E_i, E_j) and (F_i', F_j', E_i', E_j') satisfy (23) and (24),

$$(A_i + B_i^a F_i, B_i^a E_i)\text{-controlled invariance}$$

is equivalent to

$$(A_i + B_i^a F_i', B_i^a E_i')\text{-controlled invariance,}$$

by the definitions of (F_i, F_j, E_i, E_j) and (F_i', F_j', E_i', E_j'). Besides, note that the above results for characterizing mode distinguishability using controlled invariance can also be derived using the augmented system approach in [11], [12], [13], [14].

B. Controlled Mode Distinguishability

The analysis of controlled distinguishability is based on the indistinguishable subspace. Intuitively, if the control input can prevent the continuous state of system (1) from getting into the indistinguishable subspace of a pair of modes, then the two modes are controlled distinguishable. A characterization is presented in Theorem 3. First we have the following preliminary result.

Lemma 4: Suppose $x_0 \in M_{ij}$ and $u_0 \in \mathbb{R}^m$. Then there exist u_0^i and u_0^j such that

$$\chi_i(1, x_0, u_0^c, [u_0^i]) = \chi_j(1, x_0, u_0^c, [u_0^j]) \in M_{ij} \tag{27}$$

if and only if there are u^i and u^j such that

$$B_i^c u_0^c + B_i^a u^i = B_j^c u_0^c + B_j^a u^j \in M_{ij}. \tag{28}$$

Proof: Since $x_0 \in M_{ij}$, there are attack inputs $\bar{u}_0^i$ and $\bar{u}_0^j$ such that

$$A_i x_0 + B_i^a \bar{u}_0^i = A_j x_0 + B_j^a \bar{u}_0^j \in M_{ij}. \tag{29}$$

Equation (27) is equivalent to

$$A_i x_0 + B_j^c u_0^c + B_i^a u_0^i = A_j x_0 + B_j^c u_0^c + B_i^a u_0^j \in M_{ij},$$

and using (29), this is equivalent to (28) with $u_i = u_0^i - \bar{u}_0^i$ and $u_j = u_0^j - \bar{u}_0^j$. ∎

Theorem 3: Modes i and j of system (1) are not controlled distinguishable if and only if for each u_0^c, there are u_0^i and u_0^j such that (28) holds where M_{ij} is the indistinguishable subspace for modes i and j.

Proof: To demonstrate the "only if" part of the theorem, suppose that modes i and j are not controlled indistinguishable. Then, there exists $x_0 \in M_{ij}$ such that for any u_0^c, there are u_0^i and u_0^j satisfying (27). It now follows from Lemma 4 that there are u^i and u^j such that (28) holds.

To demonstrate the "if" part of the theorem, consider any x_0 in M_{ij} and any control input sequence $U_{0:\infty}^c$. We will show that there exist input sequences $U_{0:\infty}^i$ and $U_{0:\infty}^j$ such that

$$\chi_i(k, x_0, U_{0:k-1}^c, U_{0:k-1}^i) = \chi_j(k, x_0, U_{0:k-1}^c, U_{0:k-1}^j) \in M_{ij} \tag{30}$$

for all $k \geq 1$.

Consider $k = 1$. It follows from Lemma 4 that the existence of inputs u^i and u^j satisfying (28) implies the existence of inputs u_0^i and u_0^j such that (27) holds. Thus (30) holds for $k = 1$ and $U_{0:0}^q = [u_0^q]$ for $q = i, j$.

Now suppose that for some $K \geq 1$, there are sequences $U_{0:K-1}^i$ and $U_{0:K-1}^j$ such that (30) holds for $k = 1, \ldots, K$ and let

$$x_K \triangleq \chi_i(K, x_0, U_{0:K-1}^c, U_{0:K-1}^i) = \chi_j(K, x_0, U_{0:K-1}^c, U_{0:K-1}^j)$$

Since $x_K \in M_{ij}$, there are u_K^i and u_K^j such that

$$\chi_i(1, x_K, u_K^c, [u_K^i]) = \chi_j(1, x_K, u_K^c, [u_K^j]) \in M_{ij}. \tag{31}$$

Letting

$$U^q_{0:K} = [{U^q_{0:K-1}}^T \ \ {u^q_K}^T]^T \qquad q = i, j,$$

we obtain that

$$\chi_i(K+1, x_0, U^c_{0:K}, U^i_{0:K}) = \chi_j(K+1, x_0, U^c_{0:K}, U^j_{0:K}) \in M_{ij}.$$

that is, (30) holds for $k = K+1$

By induction, we obtain sequences $U^i_{0:\infty}$ and $U^j_{0:\infty}$ such that (30) holds for all $k \geq 1$. Hence modes i and j of system (1) are not controlled distinguishable ■

Corollary 5: Modes i and j of system (1) are controlled K-step distinguishable for some K if and only if they are controlled $n+1$-step distinguishable, where n is the dimension of the state space.

Proof: If $x_0 \notin M_{ij}$, then with $u^c_k \equiv 0$, it is an n-step controlled distinguishable state. Suppose $x_0 \in M_{ij}$. Since modes i and j are controlled distinguishable, it follows from Theorem 3 that there is a u_c such that (28) does not hold for any u^i and u^j. It now follows from Lemma 4 that (27) does not hold for any u^i_0 and u^j_0. Hence for each pair u^i_0, u^j_0, either $\chi_i(1, x_0, u^c_0, [u^i_0]) \neq \chi_j(1, x_0, u^c_0, [u^j_0])$ or $x_1 \in M_{ij}$ where $x_1 = \chi_i(1, x_0, u^c_0, [u^i_0]) = \chi_j(1, x_0, u^c_0, [u^j_0])$. Thus x_0 is controlled distinguishable in one step or $n+1$ steps. ■

Similar to Definition 3, we now introduce:

Definition 4 (K-step Controlled Distinguishable State): A state x_0 is called a K-step controlled distinguishable state if there is a $U^c_{0:K-1}$ such that

$$\mathcal{X}_i(K, x_0, U^c_{0:K-1}, U^i_{0:K-1}) \neq \mathcal{X}_j(K, x_0, U^c_{0:K-1}, U^j_{0:K-1}) \quad (32)$$

for any $U^i_{0:K-1}$ and any $U^j_{0:K-1}$. A state x_0 is not controlled distinguishable if it is not K-step controlled distinguishable for any K.

The proof of "if" statement of Theorem 3 implies the following result.

Corollary 6: Suppose modes i and j of system (1) are not controlled distinguishable. Then the collection of states that are not controlled distinguishable is the indistinguishable subspace M_{ij}.

C. Discerning Input Design via Quadratic Gaming

We propose the following quadratic gaming problem for obtaining a discerning input:

$$\max_{U^c_{0:n} \in \mathcal{D}} \min_{U^i_{0:n}, U^j_{0:n}} r_{x_0}(U^c_{0:n}, U^i_{0:n}, U^j_{0:n}), \quad (33)$$

where $\mathcal{D}$ is an ellipsoid without loss of generality, and

$$\begin{aligned} &r_{x_0}(U^c_{0:n}, U^i_{0:n}, U^j_{0:n}) \\ &\triangleq \left|\left|\mathcal{X}_i(n+1, x_0, U^c_{0:n}, U^i_{0:n}) - \mathcal{X}_j(n+1, x_0, U^c_{0:n}, U^j_{0:n})\right|\right| \\ &= \left|\left|\Delta\mathcal{O}_i(n+1)x_0 + \Phi^c_{ij}(n+1)U^c_{0:n} + \Phi^a_{ij}(n+1)[U^{iT}_{0:n}\ U^{jT}_{0:n}]^T\right|\right| \end{aligned}$$

where

$$\Phi^c_{ij}(n+1) \triangleq \Phi^c_i(n+1) - \Phi^c_j(n+1)$$

Here, $||\cdot||$ denotes the standard 2-norm. An optimal discerning input sequence $\hat{U}^c_{0:n}$ from (33) maximizes the distance between state trajectories of modes i and j under "worst-case" attacks.

Given system (1), suppose we can access the continuous state and we know that the discrete state is either constantly i or constantly j. Then, we can design the discerning input $\hat{U}^c_{0:n}$ using (33) and apply it to the system. Then, a mode estimate can be generated after $n+1$ times steps using the following minimum-distance criterion adapted from [19]:

$$\hat{q} \in \operatorname*{argmin}_{q \in \{i,j\}} \tilde{r}^q \quad (34)$$

where $\hat{q}$ is an estimate of the mode generated at $n+1$, $X^o_{1:n+1}$ is the sequence of observed states, and $\tilde{r}^q$ is the residual generated at $n+1$ defined by

$$\tilde{r}^q \triangleq \min_{U^q_{0:n}} \left|\left|X^o_{1:n+1} - \mathcal{X}_q(n+1, x_0, \hat{U}^c_{0:n}, U^q_{0:n})\right|\right|.$$

The following theorem reveals that the discerning input from (33) and the minimum-distance criterion (34) can jointly identify the mode.

Theorem 4: Suppose modes i and j of system (1) are controlled distinguishable, then

(i) for any $x_0, \epsilon > 0$, there exists a discerning input sequence $U^c_{0:n}$ with magnitude smaller than ϵ;

(ii) if a discerning input is designed using (33) and applied to the system, then modes i and j can be distinguished using the minimum-distance criterion within $n+1$ time steps.

Proof: To prove the first statement, for a fixed $U^c_{0:n}$, we let

$$\begin{aligned} d(U^c_{0:n}) &\triangleq \min_{U^i_{0:n}, U^j_{0:n}} r_{x_0}(U^c_{0:n}, U^i_{0:n}, U^j_{0:n}) \\ &= \left|\left|P\Big(\Delta\mathcal{O}_{ij}(n+1)x_0 + \Phi^c_{ij}(n+1)U^c_{0:n}\Big)\right|\right| \quad (35) \end{aligned}$$

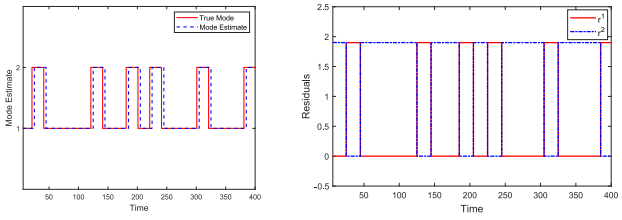
where P is the projection matrix onto the orthogonal complement of $Im\{\Phi^a_{ij}(n+1)\}$. If $x_0 \notin M_{ij}$, then $\Delta\mathcal{O}_{ij}(n+1)x_0 \notin Im\{\Phi^a_{ij}n+1)\}$ and $U^c_{0:n} = 0$ is a discerning input. If $x_0 \in M_{ij}$, then $\Delta\mathcal{O}_{ij}(n+1)x_0 \in Im\{\Phi^a_{ij}(n+1)\}$ and

$$d(U^c_{0:n}) = \left|\left|P\Phi^c_{ij}(n+1)U^c_{0:n}\right|\right| \quad (36)$$

Since x_0 is controlled distinguishable, there exists $U^c_{0:n}$ such that $\Phi^c_{ij}(n+1)U^c_{0:n} \notin Im\{\Phi^a_{ij}(n+1)\}$. It now follows from (36) that, given any $\epsilon > 0$ there is a $U^c_{0:n}$ with $\|U^c_{0:n}|| < \epsilon$ such that $d(U^c_{0:n}) > 0$. For the second statement, suppose a discerning input $U^c_{0:n}$ is applied and the true mode is mode i without loss of generality, then the residual for mode j is greater than $d(U^c_{0:n}) > 0$. ■

Remark 3: The first statement of Theorem 4 implies that the magnitude of a discerning input can be independent of either x_0 or attack inputs, which differentiates our work from [16] and [17]. The uniform boundedness of discerning input is desirable because (i) the system operator may not have the knowledge of attack inputs, and (ii) the bounded-attack-bounded-state stability of the system will not be affected.

Remark 4: It is possible to generalize the proposed quadratic gaming approach for discerning input design to the case where only sensor measurements are accessible and to the case where there is disturbance or noise. For the case where only the sensor measurements are available, we may consider alternative definitions for controlled mode distinguishability to study whether an undiscerning control input can be effectively



(a) Mode Estimate (b) Residuals

Fig. 1. (a) Time History of Mode Estimate (b) Time History of Residuals.

avoided using the output of the system, and the formulation for discerning input design can still be in the form of (33). To deal with disturbance and noise, we may either consider them as unknown-but-bounded and directly incorporate them in (33), or introduce a probability measure to them such that (33) can be modified as a stochastic gaming.

IV. Numerical Example

Consider a two-mode discrete-time switched system with the following system matrices:

$$A_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -0.1 & -0.1 & -0.1 & -0.1 \end{bmatrix}, \; B_1^a = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \; B_1^c = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -0.1 & -0.1 & -0.1 & -0.1 \end{bmatrix}, \; B_2^a = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \; B_2^c = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

It can be verified that the two modes are indistinguishable with

$$M_{12}^1 = \text{span}\{e_1, e_3, e_4\}, \quad M_{12}^2 = \text{span}\{e_1, e_4\},$$
$$M_{12}^3 = \text{span}\{e_1\}, \quad M_{12} = M_{12}^4 = \text{span}\{e_1\}, \tag{37}$$

where

$$e_1 = [1 \;\; 0 \;\; 0 \;\; 0]^T, e_3 = [0 \;\; 0 \;\; 1 \;\; 0]^T, e_4 = [0 \;\; 0 \;\; 0 \;\; 1]^T$$

We would like to note that if B_1^a and B_2^a are zero vectors, the two modes are distinguishable, and it needs exactly 4 steps to distinguish them when $x_0 = e_1$, which illustrates the non-conservativeness of our answer to Problem 1 (Note that higher dimensional examples can be constructed in the same way using companion matrices).

For the switched system given in (37), if no discerning input is applied and the initial state is in the indistinguishable subspace, then the attacker can make residuals for both modes exactly zero over time with $u_k^a = 0.1\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}x_k$, for which the minimum-distance criteria cannot provide a mode estimate. Although modes 1 and 2 are indistinguishable, they are controlled distinguishable according to Theorem 3. Therefore, the discerning input can be designed by solving the quadratic gaming problem (33). Fig. 1 shows the time histories of mode estimate and residuals when the discerning input is applied.

In the simulation, the initial state is e_1. The residuals, discerning inputs, and mode estimate are generated every 5 time steps. The magnitude of the discerning input sequence is bounded by one. From Fig. 1 (a), we see that the true mode can be identified within 10 time steps as long as the mode switching is not too frequent. From Fig. 1 (b), it is observed that the residuals are well-separated such that there is no confusion when the minimum-distance criterion is applied.

V. Conclusion

Motivated by cybersecurity issues, we have revisited mode distinguishability, characterized controlled mode distinguishability, and formulated the discerning input design problem as a quadratic gaming problem for switched linear systems with unknown inputs.

References

[1] F. Pasqualetti, F. Dorfler, and F. Bullo, "Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 110–127, Feb. 2015.

[2] V. L. Do, L. Fillatre, I. Nikiforov, and P. Willett, "Security of SCADA systems against cyber–physical attacks," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 32, no. 5, pp. 28–45, May 2017.

[3] Q. Zhu and T. Basar, "Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 46–65, Feb. 2015.

[4] K.-D. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proc. IEEE*, vol. 100, pp. 1287–1308, May 2012.

[5] P. Derler, E. A. Lee, and A. S. Vincentelli, "Modeling cyber–physical systems," *Proc. IEEE*, vol. 100, no. 1, pp. 13–28, 2011.

[6] S. Z. Yong, M. Zhu, and E. Frazzoli, "Resilient state estimation against switching attacks on stochastic cyber-physical systems," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, 2015, pp. 5162–5169.

[7] J. Hu, J. Shen, and D. Lee, "Mode-conscious stabilization of switched linear control systems against adversarial switchings," *IEEE Trans. Autom. Control*, to be published.

[8] D. Lee and J. Hu, "Stabilizability of discrete-time controlled switched linear systems," *IEEE Trans. Autom. Control*, vol. 63, no. 10, pp. 3516–3522, Oct. 2018.

[9] H. Kim, P. Guo, M. Zhu, and P. Liu, "Attack-resilient estimation of switched nonlinear cyber-physical systems," in *Proc. Amer. Control Conf. (ACC)*, 2017, pp. 4328–4333.

[10] N. Forti, G. Battistelli, L. Chisci, and B. Sinopoli, "A bayesian approach to joint attack detection and resilient state estimation," in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, 2016, pp. 1192–1198.

[11] R. Vidal, A. Chiuso, and S. Soatto, "Observability and identifiability of jump linear systems," in *Proc. 41st IEEE Conf. Decis. Control*, vol. 4, 2002, pp. 3614–3619.

[12] E. De Santis, "On location observability notions for switching systems," *Syst. Control Lett.*, vol. 60, no. 10, pp. 807–814, 2011.

[13] D. Gómez-Gutiérrez, A. Ramírez-Treviño, J. Ruiz-León, and S. Di Gennaro, "On the observability of continuous-time switched linear systems under partially unknown inputs," *IEEE Trans. Autom. Control*, vol. 57, no. 3, pp. 732–738, Mar. 2012.

[14] G. Fiore, E. De Santis, and M. D. D. Benedetto, "Secure mode distinguishability for switching systems subject to sparse attacks," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 9361–9366, 2017.

[15] L. Vu and D. Liberzon, "Invertibility of switched linear systems," *Automatica*, vol. 44, no. 4, pp. 949–958, 2008.

[16] K. M. D. Motchon, K. M. Pekpe, and J.-P. Cassar, "Robust discerning controls for the operating modes of linear switched systems subject to bounded unknown inputs," *Automatica*, vol. 96, pp. 159–165, Oct. 2018.

[17] M. Baglietto, G. Battistelli, and L. Scardovi, "Active mode observability of switching linear systems," *Automatica*, vol. 43, no. 8, pp. 1442–1449, 2007.

[18] G. Basile and G. Marro, "Controlled and conditioned invariant subspaces in linear system theory," *J. Optim. Theory Appl.*, vol. 3, no. 5, pp. 306–315, 1969.

[19] G. Battistelli, "On stabilization of switching linear systems," *Automatica*, vol. 49, no. 5, pp. 1162–1173, 2013.